History, Teaching, and Public Awareness

Research Data Management and Data Sharing for Reproducible Research—Results of a Community Survey of the German National Research Data Infrastructure Initiative Neuroscience

Carsten M. Klingner,[1,2] Michael Denker,[16] Sonja Grün,[16,17] Michael Hanke,[13,14] Steffen Oeltze-Jafra,[18,19] Frank W. Ohl,[20,21] Janina Radny,[8,9] Stefan Rotter,[15] Hansjörg Scherberger,[11,12] Alexandra Stein,[8] Thomas Wachtler,[10] Otto W. Witte,[1] and Petra Ritter[3,4,5,6,7]

[1]Hans Berger Department of Neurology, Jena University Hospital, Jena, 07747, Germany, [2]Biomagnetic Center, Jena University Hospital, Jena, 07747, Germany, [3]Berlin Institute of Health at Charité - Universitätsmedizin Berlin, Berlin, 10117, Germany, [4]Department of Neurology with Experimental Neurology, Brain Simulation Section, Charité - Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, 10117, Germany, [5]Bernstein Center for Computational Neuroscience Berlin, Berlin, 10117, Germany, [6]Einstein Center for Neuroscience Berlin, Berlin, 10117, Germany, [7]Einstein Center Digital Future, Berlin, 10117, Germany, [8]Bernstein Coordination Site, Jülich, 79104, Germany, [9]University of Freiburg, Freiburg im Breisgau, 79098, Germany, [10]Faculty of Biology, Ludwig-Maximilians-Universität München, München, 82152, Germany, [11]Deutsches Primatenzentrum GmbH – Leibniz-Institut für Primatenforschung, Göttingen, 37077, Germany, [12]Faculty of Biology and Psychology, University of Göttingen, Göttingen, 37073, Germany, [13]Institute of Neuroscience and Medicine, Brain & Behaviour (INM-7), Research Center Jülich, Jülich, 52428, Germany, [14]Institute of Systems Neuroscience, Medical Faculty, Heinrich Heine University Düsseldorf, Düsseldorf, 40225, Germany, [15]Bernstein Center Freiburg and Faculty of Biology, University of Freiburg, Freiburg im Breisgau, 79104, Germany, [16]Institute of Neuroscience and Medicine (INM-6) and Institute for Advanced Simulation (IAS-6) and JARA-Institute Brain Structure-Function Relationships (INM-10), Jülich Research Centre, Jülich, 52428, Germany, [17]Theoretical Systems Neurobiology, RWTH Aachen University, Aachen, 52074, Germany, [18]Department of Neurology, Otto von Guericke University, Magdeburg, 39120, Germany, [19]Peter L. Reichertz Institute for Medical Informatics, Hannover Medical School, Hannover, 30625, Germany, [20]Leibniz Institute for Neurobiology (LIN), Magdeburg, 38118, Germany, and [21]Center for Behavioral Brain Science (CBBS), Magdeburg, 39106, Germany

Abstract

Science is changing: the volume and complexity of data are increasing, the number of studies is growing and the goal of achieving reproducible results requires new solutions for scientific data management. In the field of neuroscience, the German National Research Data Infrastructure (NFDI-Neuro) initiative aims to develop sustainable solutions for research data management (RDM). To obtain an understanding of the present RDM situation in the

Significance Statement

A comprehensive survey among the neuroscience community in Germany determined the current needs, challenges, and opinions with respect to standardized research data management (RDM). The Neuroscience community perceives a lack of standards for data and metadata, a lack of provenance tracking and versioning of data, a lack of protected digital research infrastructure for sensitive data and a lack of education and resources for proper RDM. However, an overwhelming majority of community members indicated that they would be willing to share their data with other researchers and are interested to increase their RDM skills. Thus, the survey results suggest that training, the provision of standards, tools, infrastructure, and resources for RDM holds the potential to significantly facilitate reproducible research in neuroscience.

neuroscience community, NFDI-Neuro conducted a comprehensive survey among the neuroscience community. Here, we report and analyze the results of the survey. We focused the survey and our analysis on current needs, challenges, and opinions about RDM. The German neuroscience community perceives barriers with respect to RDM and data sharing mainly linked to (1) lack of data and metadata standards, (2) lack of community adopted provenance tracking methods, (3) lack of secure and privacy preserving research infrastructure for sensitive data, (4) lack of RDM literacy, and (5) lack of resources (time, personnel, money) for proper RDM. However, an overwhelming majority of community members (91%) indicated that they would be willing to share their data with other researchers and are interested to increase their RDM skills. Taking advantage of this willingness and overcoming the existing barriers requires the systematic development of standards, tools, and infrastructure, the provision of training, education, and support, as well as additional resources for RDM to the research community and a constant dialogue with relevant stakeholders including policy makers to leverage of a culture change through adapted incentivization and regulation.

Key words: research data infrastructure; data sharing; metadata; survey; community; information security and privacy

Introduction

Annual brain health costs exceed €800 billion in Europe (DiLuca and Olesen, 2014). Many factors contribute to the difficulty of developing effective treatments for brain diseases. These include the gaps in knowledge about the precise changes and biological processes in the brain that cause a disease, and the long time needed to observe whether an investigational treatment affects disease progression. Many studies have been collecting cohort datasets in patients with brain disease to better understand the mechanistic basis of the disease, qualify diagnostic and monitoring biomarkers, and test drugs. Given that most studies are typically limited in their range of assessments, there is tremendous value in combining and integrating the resulting data (Milham et al., 2018). Multimodal data across different studies can further been used to construct integrated *in silico* models of the brain and multiscale potential targets for

Copyright © 2023 Klingner et al.

multilevel interventions. However, the proportion of scientific data that is actually openly shared within the neuroscientific community remains low (Watson, 2022). The lack of sharing properly annotated data and tools contributes to the poor reproducibility of research results, known as "the reproducibility crisis," that hinders the growth of knowledge and innovation on the one hand and leads to inefficient use of resources on the other hand (Baker, 2016; Stodden et al., 2016; Poldrack et al., 2019; Crook et al., 2020; Loss et al., 2021; Niso et al., 2022).

The German National Research Data Infrastructure Initiative (NFDI) implemented by the German Research Foundation (DFG) will provide up to €85 million per year over the course of 10 years (https://dfg.de/nfdi) to foster research data management (RDM) across all research domains in Germany. RDM describes the organization, storage, preservation, and sharing of scientific data. This includes the day-to-day management of research data during the lifetime of a research project and the long-term usability of these data through the FAIR principles (findable, accessible, interoperable, and reusable). NFDI comprises domain-specific consortia across all science disciplines. In the field of neuroscience, the initiative NFDI Neuroscience (NFDI-Neuro; https://nfdi-neuro.de) initiative has started to closely interact with the neuroscience community to overcome the challenges in RDM (Denker et al., 2021a,b; Hanke et al., 2021; Klingner et al., 2021; Wachtler et al., 2021).

The NFDI-Neuro initiative is aligned to several international programs such as the WHO Global strategy on digital health, the European Health Data Space, the European Interoperability Framework (EIF) and the Digital Europe Program (https://digital-strategy.ec.europa.eu/en/activities/digital-programme) by addressing topics such as interoperability, fair digital objects, artificial intelligence, and cybersecurity. A goal of NFDI-Neuro is to foster the reproducibility of research and to leverage computational neuroscience as data integrating discipline that transforms data into knowledge and understanding.

To obtain a comprehensive understanding of the present RDM situation in the neuroscience community, NFDI-Neuro conducted a community survey to investigate what among neuroscience community is perceived as the largest obstacles and most pressing needs with respect to
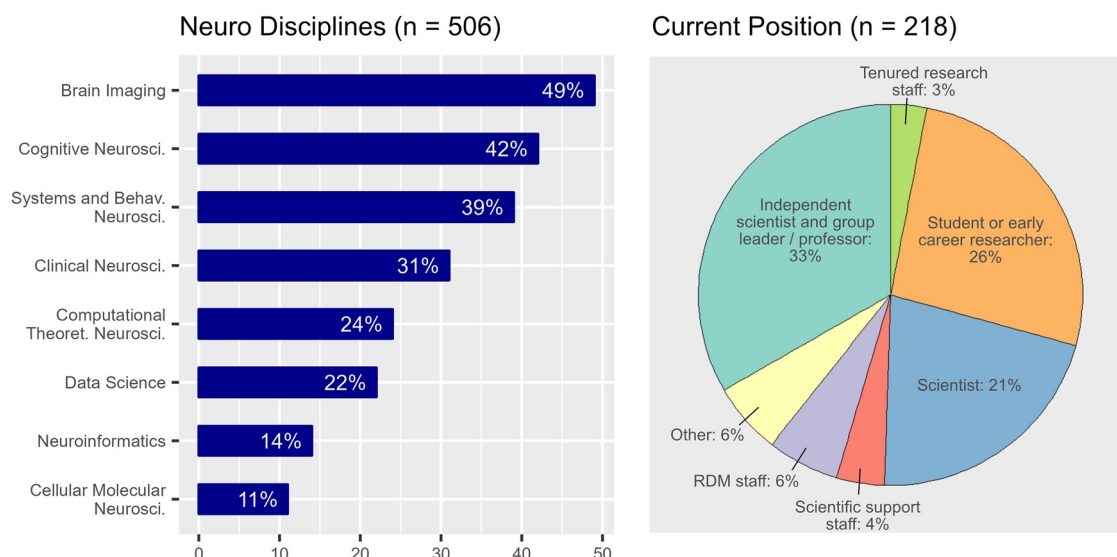
Figure 1. Distribution of neuroscience subdisciplines (multiple answers allowed, left), professional position of participants (right).

RDM, and how members of the community self-assess their present proficiency in RDM topics.

Materials and Methods

The NFDI-Neuro community survey was developed based on a previous survey by the partner consortium NFDI4Bioimage (https://nfdi4bioimage.de/). It was adapted by the NFDI-Neuro team to address questions specific to the neuroscience research domain. The NFDI-Neuro survey comprised 20 sets of questions, where each set contained one or multiple questions. Counting all questions yields a total of 114 questions presented to each survey participant. The time required for answering all questions was 10–15 min. We used the tool LimeSurvey (https://www.limesurvey.org/de/) and conducted the survey online in compliance with the EU General Data Protection Regulations (GDPR). The online questionnaire was made available for two months between September 1, 2021 and November 1, 2021 via the website of the NFDI-Neuro initiative (https://nfdi-neuro.de/). We announced the survey via several channels, including E-mail lists of the German neuroscience communities, such as the German Society for Clinical Neurophysiology, the German Neuroscience Society, and the Bernstein Network Computational Neuroscience, as well as the NFDI-Neuro community mailing list and Twitter channel (https://twitter.com/NFDI_Neuro). In total, 218 individuals of either sex participated in the online survey. Of those, 85 participants did not answer all questions. We included in our analysis all given answers, including those of the incomplete questionnaires. For the data analysis and the generation of the figures, we used the software package R [version 4.1.2 ("Bird Hippie")]. The survey and related collected data, as well as all analysis scripts are available publicly (https://doi.org/10.12751/g-node.w5h68v).

Data availability

Raw data of the survey have been published in a research data repository (https://doi.org/10.12751/g-node.w5h68v).

Results

In the following, we present the main results of the survey. A visual representation of the answers to each question can be found in the supplementary material.

Participants represent a broad range of neuroscience disciplines

Most respondents work at a public university or government research institution (71%), while 19% work at a nonprofit research institute and 5% work at a private company. The distribution of professional positions of the survey participants shows a tendency toward higher positions in the scientific hierarchy, with 73 (33%) "Independent scientist and group leader/professor," 46 (21%) "Scientists," 56 (26%) "Student or early career researcher," 14 (6%) "Research data management focused staff," 6 (3%) "Tenured research staff," 9 (4%) "Scientific support staff," 14 (6%) "Other" (Fig. 1). The participants cover a wide range of neuroscience subdisciplines (Fig. 1; selection of multiple choices possible) led by brain imaging (106, 49%) followed by cognitive neuroscience (92, 42%), systems and behavioral neuroscience (84, 39%), clinical neuroscience (67, 31%), computational/theoretical neuroscience (53, 24%), data science (48, 22%), neuroinformatics (31, 14%), and cellular/molecular neuroscience (25, 11%).

A considerable amount of research data is not yet being shared

A total of 114 (79%) of all participants indicated that they share data within their institution. 95 (66%) share their data with external collaborators while only 65 (45%) share data publicly, that is they made datasets openly discoverable via repositories (at least one dataset). Only 13 participants (9%) had never shared any data yet (Fig. 2). A primary objective of the NFDI initiative is to improve the reuse of research data. In this context, we explored the potential availability of neuroscience data that is not yet
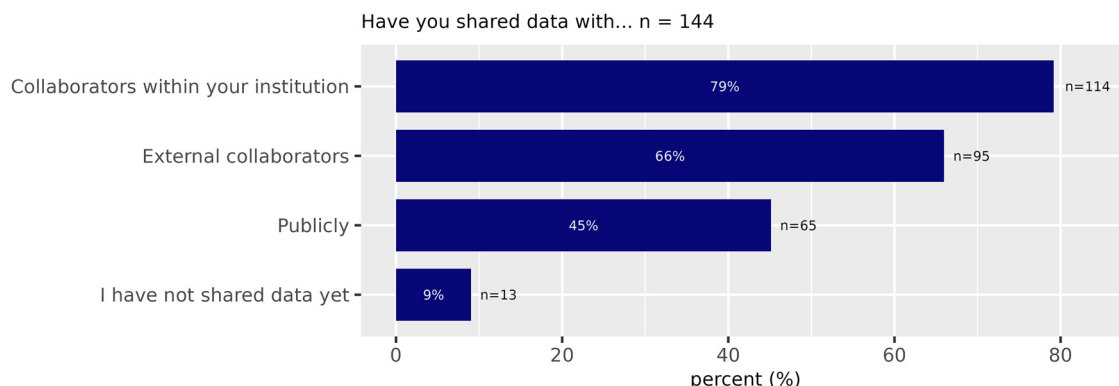
Have you shared data with... n = 144

Figure 2. Data sharing (Survey Question 7).

shared publicly but is considered of general interest. We asked whether the participants own data of potential interest to other scientists for reuse (Fig. 3). According to the responses, 84 (67%) of the participants have valuable datasets available that would be useful for further exploitation, but only 20 (22%) of those participants make these data available for reuse. 76 (84%) of all participants with at least one dataset believed that other researchers could answer their research questions by reusing data from their research. However, even for this subgroup of scientists that think their data are valuable to others, 48% have never publicly (via public repositories) shared any of their data.

Own data management skills are commonly seen as not being high

Research data management skills are essential for preparing, analyzing, and publicly sharing data. 43% of responders disagreed with the statement "Overall I am highly knowledgeable about research data management in my research field" (Fig. 4). Only 33% of the survey participants thought that they have proficiency in research data management. Only 36% thought they know which research data management methods are available, and 36% thought they are "highly knowledgeable about research data management." Interestingly, 58% of all respondents nevertheless agreed or rather agreed that they "can handle their research data according to community

standards." It is unlikely that this apparent discrepancy could be explained by the availability of data research managers who assist with data handling because only 19% of participants indicated that they have dedicated personnel with research data management or data curation expertise in their labs.

We further investigated whether there is a dependency between public data sharing and the self-perception of competence regarding RDM. From the reported self-assessments, only the statement "I think that I can handle research data according to community standards" (Fig. 4) showed a strong connection to the response that data are shared openly (Fig. 2). Participants agreeing to this statement were six times more likely to share data publicly than those who were disagreeing. Self-assessed high competence in the other RDM capabilities correlated with reported data sharing as well, but to lower degrees ("I know which research data management methods are available" 1.2-fold, "Overall, I am highly knowledgeable about research data management in my research field" 1.4-fold, "I have proficiency in RDM" 1.75-fold). Thus, in summary, the higher the level of RDM knowledge, the higher the level of data sharing.

Tools and standards for RDM are not yet widely used

While the responses indicated that standard tools for data processing and analysis are widely used, the use of
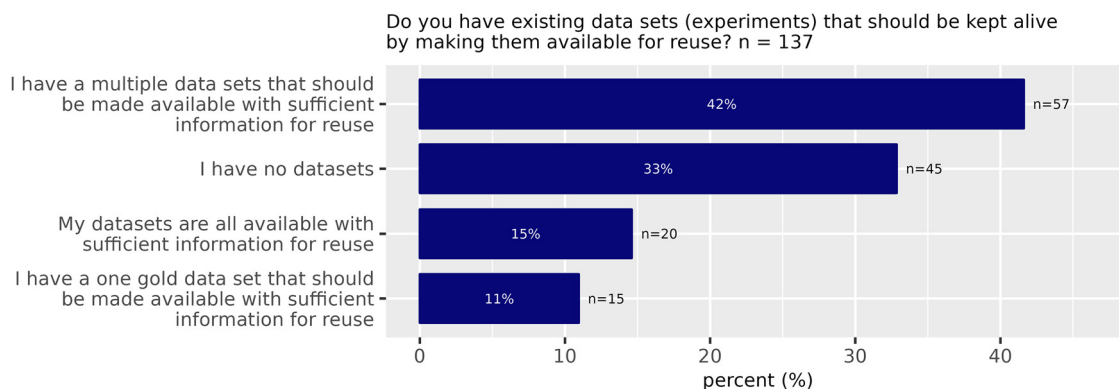
Do you have existing data sets (experiments) that should be kept alive by making them available for reuse? n = 137

Figure 3. Existing datasets (Survey Question 8).

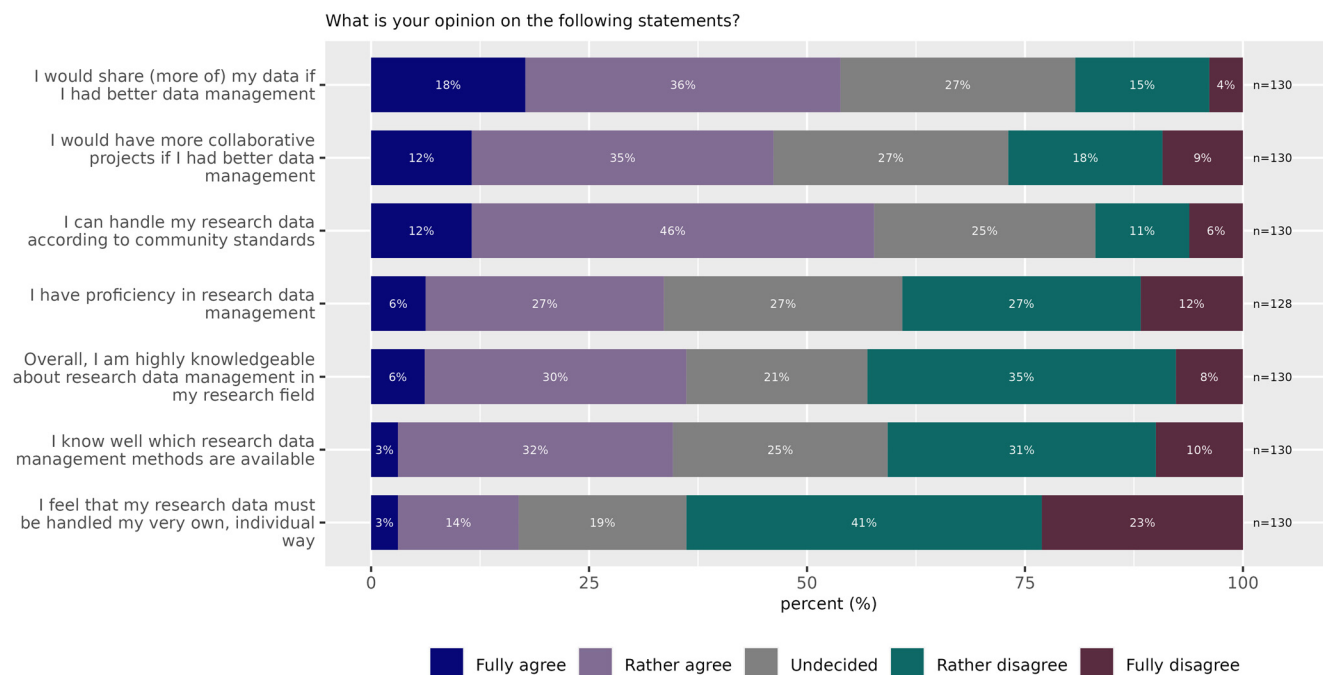What is your opinion on the following statements?

Figure 4. Perceived obstacles for RDM and data sharing (Survey Question 14).

standard RDM tools for data sharing was reported to significantly lower extent (sharing data openly, metadata collection and management, provenance tracking; Fig. 5).

Scientists who use more tools or standards are more likely to share their data

In the group that indicated not sharing their data publicly, only 33% indicated using tools or standards, while this proportion was 54% in the group of data sharers. A possible explanation could be that scientists who work a lot with standard tools find it easier to present their data in a shareable form and have a higher digital literacy required for the public sharing of data. Alternatively, the motivation to share data may be a strong driver to adopt standard methods. Respondents who indicated sharing their data publicly were 42% more likely to use standard tools "mostly" in their daily work compared with respondents who did not share their data publicly.

Perceived obstacles for research data management and sharing

Reluctancy to share data publicly because the data ownership or intellectual property might be violated was indicated by 20% of respondents. Interestingly, 37% of participants reported that they do not know whether their institutional policy allows uploading data to a public repository, while only 9% were confident that their institutions do not support this.

Further, 58% were not sure whether they own the rights to upload data from their own research project; 48% reported seeing legal aspects as significant hurdles for public repository usage. These answers indicate that substantial uncertainties about legal issues regarding data sharing exist in the research community. This is confirmed by only 18% indicating that legal aspects were not perceived as significant hurdles for public repository usage.

Only 29% of participants thought there is sufficient guidance for choosing an appropriate repository for their data; 63% believed that there is a lack of expertise and human resources to deposit data in a repository; 45% thought that the technical hurdles are too high to upload data to a repository.

Eighty-three percent of respondents indicated that their research data did not have to be handled in an individual way not easily compatible with existing standards, tools, or guidelines (Fig. 4). The lack of professional data management was reported as a problem. A total of 70 (54%) participants responded that they would share more of their data if they had better RDM, while only 27% indicated that better RDM would not increase the amount of their own data to share.

A total of 70% of those respondents who had previously prepared data for publication and reuse indicated that the time they need to ready a dataset requires more than a day, and 26% need even more than a week (Fig. 6).

Accordingly, 60% thought that there is lack of time to deposit data in a repository. In comparison, only 23% did not believe that time is a problem for depositing data in a public repository.

Questioned for the most pressing issues hindering research data management and public data sharing, there was a strong consensus. Close to 70% of respondents rated the following problems as one of the top three:
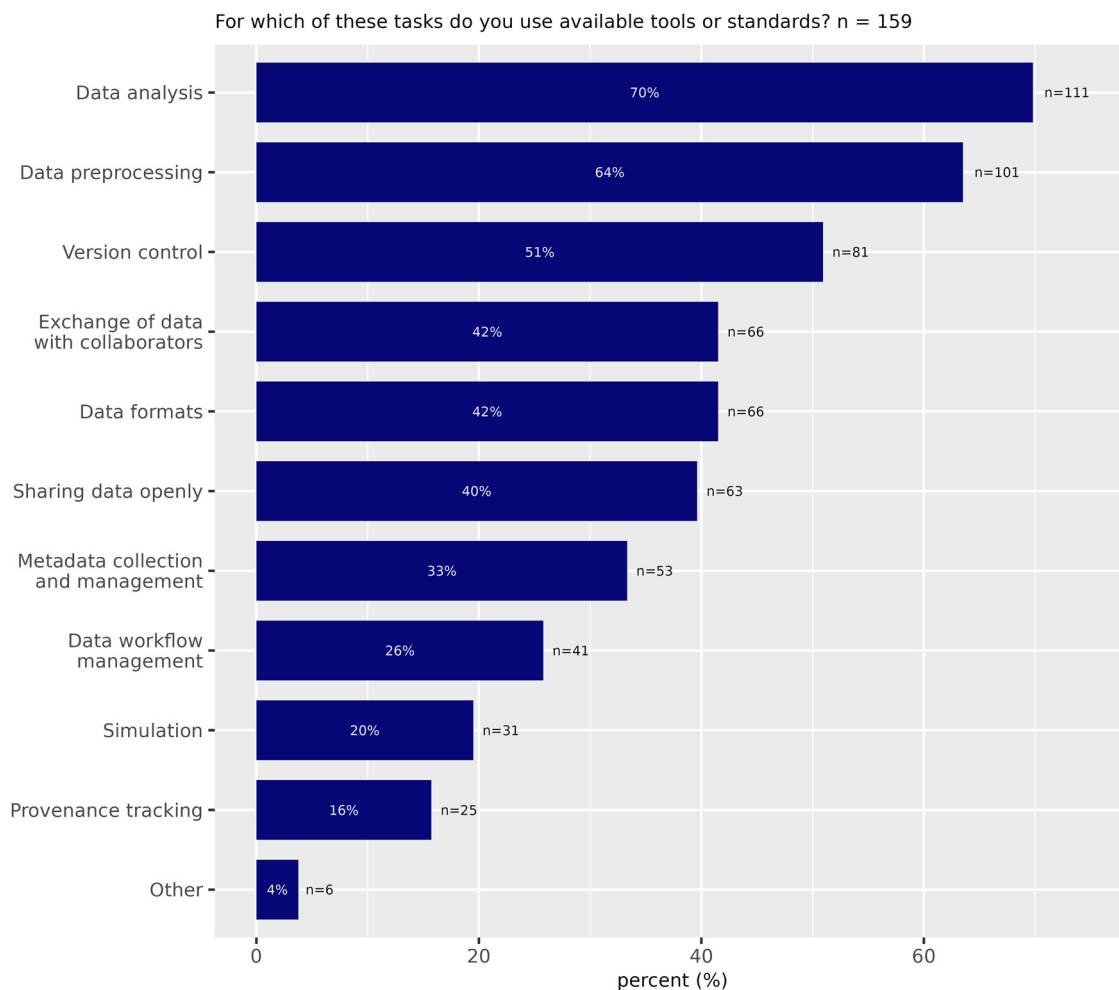
For which of these tasks do you use available tools or standards? n = 159

Task	Percent	n
Data analysis	70%	n=111
Data preprocessing	64%	n=101
Version control	51%	n=81
Exchange of data with collaborators	42%	n=66
Data formats	42%	n=66
Sharing data openly	40%	n=63
Metadata collection and management	33%	n=53
Data workflow management	26%	n=41
Simulation	20%	n=31
Provenance tracking	16%	n=25
Other	4%	n=6

percent (%)

Figure 5. Use of existing tools and standards for different research data management activities (Survey Question 5).

- "Inappropriately documented custom code in non-re-producible computational environments"
- "Poor standardization of metadata and derived data"

Other concerns were indicated similarly often but were rated with lower importance compared with the top two problems:

- "Lack of automatic data quality control"

How much time do you currently need to ready a dataset from your lab for publication and re-use? n = 128

Time	Percent	n
I don't know	34%	n=44
More than a week	26%	n=33
Less then a week	20%	n=26
A day	11%	n=14
Few hours or less	9%	n=11
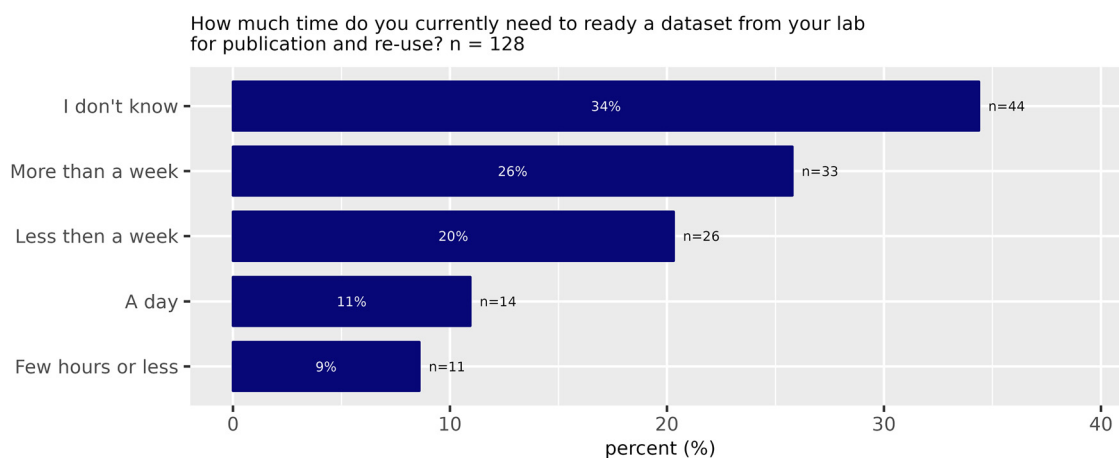
percent (%)

Figure 6. Required time to prepare a dataset for publication and reuse (Survey Question 18).

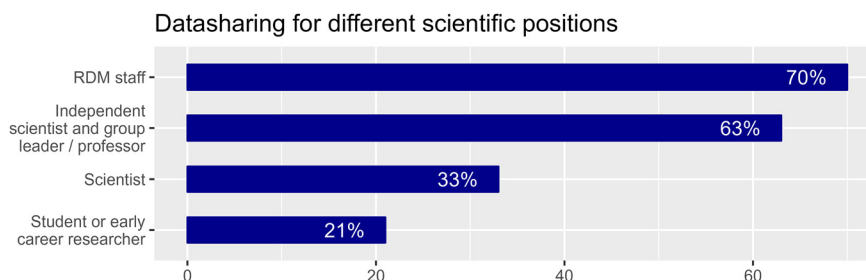## Datasharing for different scientific positions

Figure 7. Percentage of respondents that have at least one dataset shared publicly shown separately according to their scientific position.

- "Harmonization and fusion of data from multiple sites and/or studies"
- "No standardized support for concerted study data and metadata extraction from multiple devices and data linking"
- "Data security issues in data exchange with other institution"

In addition, the responses suggested that general knowledge about methods and tools of research data management is lacking. Only 34% of participants indicate that they think they know which RDM methods are available.

Factors promoting public data sharing

To identify factors that promote public data sharing, we analyzed separately the answers of the participants who had already shared their data in public repositories ($n = 65$). For this analysis we excluded responses from participants who had indicated that they had never shared a single dataset publicly. The proportions of the different academic groups among scientists who have shared data publicly varied considerably: responses to the question whether data are shared depended on the position and experience of the person managing the data (Fig. 7).

Interestingly, whether dedicated personnel with RDM or data curation expertise is available seems not to affect whether data are publicly shared. Public data sharing was reported by 56% of those indicating that dedicated RDM personnel is available and by 54% of those indicating that it was not available.

We analyzed the dependence between the willingness to share data and the scientific subdomain of the respective researcher. We found a relatively high degree of data sharing indicated by scientists in the subdomain of neuroinformatics (58%), and a relatively low degree (36%) of data sharing in clinical neuroscience (Fig. 8).

Discussion

It is generally recommended and expected by journals and founding agencies that scientific data be shared to improve collaboration, transparency, and reproducibility in science. However, less than half (45%) of the participants of the current survey stated they had made at least one dataset publicly available. While this suggests that data sharing is possible in principle, it remains unclear to what extent this 45% of scientists share the data they
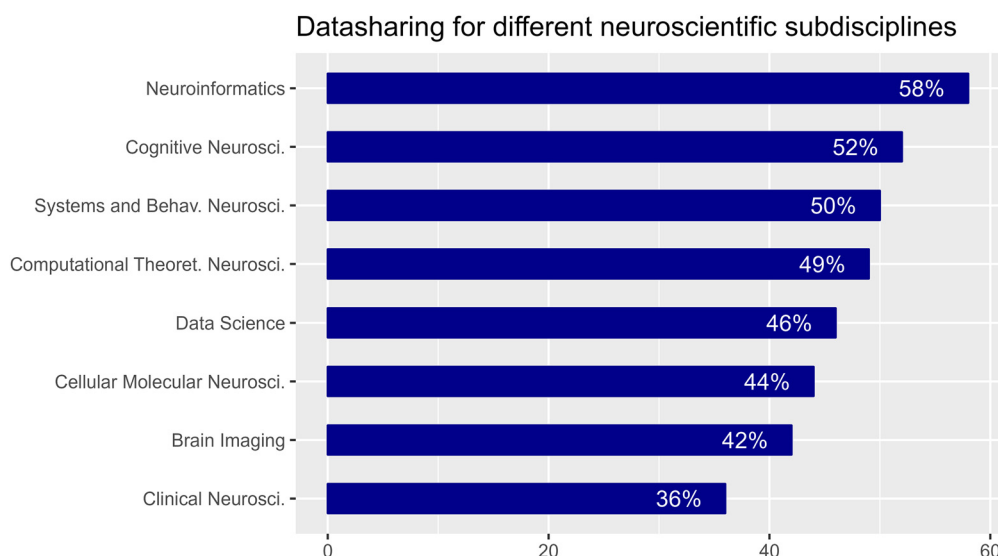
Datasharing for different neuroscientific subdisciplines

Figure 8. Percentage of respondents who have at least one dataset shared publicly, separated according to their scientific subdomain.

collect. In any case, most participants have never shared data publicly. Although the survey was conducted anonymously, there is nevertheless a tendency to answer according to social desirability or even to give an answer bias in the direction that the respondent himself would like to see. Therefore, the proportion of scientists who did not share data could be even higher. In addition, it remains unclear how much of the data they collected they shared, as a single publicly shared dataset is enough to be counted in this group. The percentage of collected data that is publicly shared remains therefore unclear but is certainly well below 45%. This result is consistent with a previous survey on open science practices in functional neuroimaging which reported that 34% of their participants have never shared their raw neuroimaging data (Paret et al., 2021). This similarity could also be caused by the fact that half of respondents in our survey are engaged in neuroimaging. Some other domains might be underrepresented like cellular molecular neuroscience (11%) or neuroinformatics (14%). The sample size of the survey is comparable to other surveys in this area and similarity of results indicate representativeness (Niso et al., 2022). The low rate of shared data can be explained by the fact that scientists do not want to or cannot share the data or that there are at least barriers that ultimately lead to the data not being shared. Respondents to our survey showed no fundamental objection to sharing and reusing scientific data. Yet the data-collecting scientist may fear specific disadvantages from sharing data, e.g., other scientists can specialize in refuting study results and make their mark at the expense of the scientists collecting the data (Longo and Drazen, 2016). Other authors argue that sharing data are worthwhile even for the "most avaricious and self-interested scientist" and leads to an improvement in their own scientific productivity and career advancement (Hunt, 2019).

Even if there is no fundamental resistance to sharing data, the results indicate that there are barriers to sharing data. In the survey, we specifically addressed respondent's perceptions of various possible barriers. Reluctancy to share data publicly because the data ownership or intellectual property might be violated was indicated by 20% of respondents. Interestingly, more than one-third of participants reported that they do not know whether their institutional policy allows uploading data to a public repository, while only one out of ten was confident that his/her institution do not support this.

Further, more than half of all respondents were not sure whether they own the rights to upload data from their own research project. A total of 48% reported seeing legal aspects as significant hurdles for public repository usage. These answers indicate that substantial uncertainties about legal issues regarding data sharing exist in the research community. This is confirmed by only a minority (18%) state that legal aspects were not perceived as significant hurdles for public repository usage. This general view is consistent with a recent review of the spectrum of data sharing policies in neuroimaging data repositories (Jwa and Poldrack, 2022). The authors highlighted the complexity of ethical and legal issues related to neuroimaging data in the United States, which depend on several factors, such as the sensitivity of the data, whether a federal agency is involved, or whether stakeholders still want to retain some control over the shared data. Because of the complexity of the solutions to the constraints and requirements, choosing the right solution requires expert knowledge, and the increasing threats posed by new technologies to the privacy of shared data lead the authors to propose a legal ban on the malicious use of neuroscience data (Jwa and Poldrack, 2022). In line with this complexity, only 29% of participants thought there is sufficient guidance for choosing an appropriate repository for their data; 63% believed that there is a lack of expertise and human resources to deposit data in a repository. However, beyond legal and ethical issues, 45% thought that also the technical hurdles are too high to upload data to a repository.

These answers do not indicate the lack of a specific solution that is urgently needed, but rather that the problem lies in the difficulty of finding and applying the right solution for the individual case.

Insufficient incentives to spend the time needed for RDM

Barriers to sharing data appear also in the requirement of data preparation, which can be a time-consuming process. A majority (70%) of those respondents who had previously prepared data for publication and reuse indicated that the time they need to ready a dataset requires more than a day, and 26% need even more than a week (Fig. 6). Accordingly, 60% thought that there is lack of time to deposit data in a repository. In comparison, only a quarter (23%) did not believe that time is a problem for depositing data in a public repository. In other words, nearly half (40%) of scientists state it takes them less than a week to prepare a dataset for publication and reuse. The fact that the majority (60%) nevertheless think there is a lack of time suggests that scientists do not think the time investment is worth it. The lack of time results from competing demands, each of which requires time. Given that scientists do science, this result suggests that preparing data for publication and reuse competes with the act of doing science, which itself is not always perceived as a central part of science.

In any case, however, the reports of lack of time indicate that scientists perceive other tasks as more important than data sharing. Accordingly, this reflects an insufficient incentive to spend the time needed to prepare data for publication and reuse. On the one hand, the incentive can be increased but also the time required can be reduced to make the ratio of data sharing more advantageous. The time required can be reduced through the use of tools and the competence in research data management can be increased.

Need for training and education

The fact that two-thirds of the participants do not consider themselves competent in this area and think they do

not know which research data management methods are available shows that there is considerable potential for optimization here. Training in this regard is particularly important as our analysis demonstrated a strong correlation between research data management skills and the amount of data sharing. Research data management skills also saving time which is perceived as a major barrier to data sharing. Training and education should also convey the value of data sharing to the scientists themselves. In a recent Nature RDM survey (Checklists Work to Improve Science, 2018), 58% of participants think that the researchers hold the key role for improving the reproducibility of research and 91% see them among the top three stakeholders to achieve this – thus being in a leading role ahead of laboratory heads, publishers, funders, department heads and professional societies who also were among the choices. This is in great alignment with what we experience in our work within RDM neuroscience projects [NFDI-Neuro, INCF, EBRAINS, central informatics projects of collaborative research centers funded by DFG (known as "INF projects"): "It is the researchers themselves who are required to do the RDM and to co-develop RDM tools - and hence require training to obtain RDM literacy." "Reproducibility is like brushing your teeth. Once you learnt it, it becomes a habit" (Irakli Loladze in Baker, 2016)]. NFDI-Neuro aims to bring RDM to the individual labs, via several mechanisms including the establishment of transfer teams, working groups and massive training offers.

Summary of perceived needs based on the survey results

The results of this survey indicate the following five task areas for improving RDM in Neuroscience:

1. Training, education, and networking activities.
2. Data and metadata standards, to advance and disseminate existing standards.
3. Provenance and workflows to advance and disseminate solutions for data lineage and digital reproducible workflows.
4. Infrastructure and service, for data management and processing, including for sensitive data with Cloud and HPC resources.
5. Modelling and big data analytics, for collecting RDM requirements from the perspective of the secondary data user community.

Limitation

Participants in this survey were not given any further information about the meaning of each term. The survey was answered as the participants understood the terms. Different levels of knowledge and also different subdisciplines could have an influence on the understanding of the terminology which in turn could affect the interpretation of the results. The subdisciplines of neuroscience were not evenly distributed among the respondents. The results of this survey can therefore not be generalized to all subdisciplines. The positions of the respondents showed a tendency toward more senior staff participating in the survey compared with the assumed distribution of positions in neuroscience research.

Conclusion

With the present survey, we identified various challenges in RDM in the neuroscience community. We found that the community perceives significant deficits with respect to transparent and reproducible data handling, annotation and sharing. According to the survey results, researchers with more experience and knowledge in RDM are more likely to share data for secondary use by their colleagues.

In summary:

- Only one-third of neuroscientists think they have proficiency in RDM.
- Less than a quarter of the research teams have RDM staff.
- More than a third do not know whether institutional policies allow loading data to a repository.
- Two-thirds are not sure they own rights for uploading data to public repositories.
- Half of the researchers see legal hurdles for data sharing.
- Forty percent of those researchers who have previously prepared data for publication and reuse say that the time they need to ready a dataset requires more than a week.
- Sixty percent think there is a lack of time to deposit data in a repository.
- Only one-third think they know which RDM methods are available.

We are encouraged by the fact that only a minority of one-fifth of respondents in the neuroscience community are not inclined to share data for reuse and that literacy in the usage of tools and standards increases the frequency of data sharing. Thus, the survey results suggest that training, the provision of properly secure and protected research infrastructure, tools, standards, and additional resources for RDM are promising approaches to leverage RDM and foster reproducible and efficient research practices in neuroscience. NFDI-Neuro will deliver on these topics. Therefore, we are convinced that we are addressing with NFDI Neuro the most pressing needs of our community. Our initiative has contributed significantly to several of the crosscutting goals of NFDI in the past. NFDI-Neuro plans to advance these operational solutions and to transfer them to an increasing number of labs of the German and international science community.

References

Baker M (2016) 1,500 scientists lift the lid on reproducibility. Nature 533:452–454.

Checklists Work to Improve Science (2018) Nature 556:273–274.

Crook SM, Davison AP, McDougal RA, Plesser HE (2020) Editorial: reproducibility and rigour in computational neuroscience. Front Neuroinform 14:23.

Denker M, Grün S, Wachtler T, Scherberger H (2021a) Reproducibility and efficiency in handling complex neurophysiological data. Neuroforum 27:27–34.

Denker M, Stein A, Wachtler T (2021b) Editorial. Neuroforum 27:1–2.

DiLuca M, Olesen J (2014) The cost of brain diseases: a burden or a challenge? Neuron 82:1205–1208.

Hanke M, Pestilli F, Wagner AS, Markiewicz CJ, Poline JB, Halchenko YO (2021) In defense of decentralized research data management. Neuroforum 27:17–25.

Hunt LT (2019) The life-changing magic of sharing your data. Nat Hum Behav 3:312–315.

Jwa AS, Poldrack RA (2022) The spectrum of data sharing policies in neuroimaging data repositories. Hum Brain Mapp 43:2707–2721.

Klingner CM, Ritter P, Brodoehl S, Gaser C, Scherag A, Güllmar D, Rosenow F, Ziemann U, Witte OW (2021) Research data management in clinical neuroscience: the national research data infrastructure initiative. Neuroforum 27:35–43.

Longo DL, Drazen JM (2016) Data sharing. N Engl J Med 374:276–277.

Loss CM, Melleu FF, Domingues K, Lino-de-Oliveira C, Viola GG (2021) Combining animal welfare with experimental rigor to improve reproducibility in behavioral neuroscience. Front Behav Neurosci 15:763428.

Milham MP, Craddock RC, Son JJ, Fleischmann M, Clucas J, Xu H, Koo B, Krishnakumar A, Biswal BB, Castellanos FX, Colcombe S, Di Martino A, Zuo XN, Klein A (2018) Assessment of the impact of shared brain imaging data on the scientific literature. Nat Commun 9:2818.

Niso G, et al. (2022) Open and reproducible neuroimaging: from study inception to publication. Neuroimage 12:263–273.

Paret C, Unverhau N, Feingold F, Poldrack RA, Stirner M, Schmahl C, Sicorello M (2021) Survey on open science practices in functional neuroimaging. NeuroImage 15:257–275.

Poldrack RA, Feingold F, Frank MJ, Gleeson P, de Hollander G, Huys QJM, Love BC, Markiewicz CJ, Moran R, Ritter P, Rogers TT, Turner BM, Yarkoni T, Zhan M, Cohen JD (2019) The importance of standards for sharing of computational models and data. Comput Brain Behav 2:229–232.

Stodden V, McNutt M, Bailey DH, Deelman E, Gil Y, Hanson B, Heroux MA, Ioannidis JPA, Taufer M (2016) Enhancing reproducibility for computational methods. Science 354:1240–1241.

Wachtler T, Bauer P, Denker M, Grün S, Hanke M, Klein J, Oeltze-Jafra S, Ritter P, Rotter S, Scherberger H, Stein A, Witte OW (2021) NFDI-Neuro: building a community for neuroscience research data management in Germany. Neuroforum 27:3–15.

Watson C (2022) Many researchers say they'll share data—But don't. Nature 606:853–853.